end-of-paragraph markers to the text, thereby producing the word grouping data (Figure 2A:46) which comprises sentence markers 86 and paragraph markers 87. The segmentation and normalisation process 80 is conventional, a fuller description of it can be found in Edgington M et al: 'Overview of current text-to-speech techniques part 1 – 'text and linguistic analysis', BT Technology Journal, Volume 14, No. 1, pp 68-83 (January 1996). The disclosure of that paper (hereinafter referred to as part 1 of the BTTJ article) is hereby incorporated herein by reference.

The computer is then controlled by the program to run a pronunciation and tagging process 90 which converts the expanded text file 88 to an unresolved phonetic transcription file 92 and adds tags 93 to words indicating their syntactic characteristics (or a plurality of possible syntactic characteristics). The process 90 makes use of the lexicon 44 which outputs possible word tags 93 and corresponding phonetic transcriptions of input words. The phonetic transcription 92 is unresolved to the extent that some words (e.g. 'live') are pronounced differently when playing different roles in a sentence. Again, the pronunciation process is conventional - more details are to be found in part 1 of the BTTJ article.

The program then causes the computer to run a conventional parsing process 94. A more detailed description of the parsing process can be found in part 1 of the BTTJ article.

The parsing process 94 begins with a stochastic tagging procedure which resolves the syntactic characteristic associated with each one of the words for which the pronunciation and tagging process 90 has given a plurality of possible syntactic characteristics. The unresolved word tags data 93 is thereby turned into word tags data 95. Once that has been done, the correct pronunciation of the word is identified to form phonetic transcription data 97. In a conventional manner, the parsing process 94 then assigns syntactic labels 96 to groups of words.

To give an example, if the sentence 'Similarly Britain became popular after a rumour got about that Mrs Thatcher had declared open house.' were to be input to the text-to-speech synthesiser, then the output from the parsing process 94 would be:

SENTSTART <ADV Similarly_RR ADV> ,_, (NR Britain_NP1 NR) [VG became_VVD VG] <ADJ popular_JJ ADJ> [pp after_ICS (NR a_AT1 rumour_NN1 NR) pp] [VG got_VVD about_RP VG] that_CST (NR Mrs_NNSB1 Thatcher_NP1 NR) [VG had_VHD declared_VVN VG] (NR open_JJ house_NNL1 NR) SENTEND ._.

Where **SENTSTART** and **SENTEND** represent the sentence markers 86, _RR, _NP1 etc. represent the word tag data 95, and **<ADV ..... ADV>, (NR ..... NR)** etc. represent the syntactic groups 96. The meanings of the word tags used in this description will be understood by those skilled in the art – a subset of the word tags used is given in Table 1 below, a full list can be found in Garside, R., Leech, G. and Sampson, G. eds 'The Computation Analysis of English : A Corpus based Approach', Longman (1987).

Word Tag	Definition
( ) , - . ... : ; ?	Punctuation
AT1	singular article: *a, every*
CST	*that* as conjunction
DA1	singular after-determiner: *little, much*
DDQ	'wh-' determiner without '-ever': *what, which*
ICS	preposition-conjunction of time: *after, before, since*
IO	*of* as preposition
JJ	general adjective
NN1	singular common noun: *book, girl*
NNL1	singular locative noun: *island, Street*
NNS1	singular titular noun: *Mrs, President*
NP1	singular proper noun: *London, Frederick*
PPH1	*it*
RP	prepositional adverb which is also particle

RR	general adverb
RRQ	non-degree 'wh-adverb' without '-ever': *where, when, why*
TO	infinitive marker *to*
UH	interjection: *hello, no*
VBO	base form *be*
VBDR	imperfective indicative *were*
VBDZ	*was*
VBG	*being*
VBM	*am,'m*
VBN	*been*
VBR	*are, 're*
VBZ	*is, 's*
VDO	base form *do*
VDD	*did*
VDG	*doing*
VDN	*done*
VDZ	*does*
VHO	base form *have*
VHD	*had, 'd* (preterite)
VVD	lexical verb, preterite: *ate, requested*
VVG	'-ing' present participle of lexical verb: *giving*
VVN	past participle of lexical verb: *given*

Table 1

Next, in chunking process 98, the program controls the computer to label 'chunks' in the input sentence. In the present embodiment, the syntactic groups shown in Table 2 below are identified as chunks.

TAG	Description	Example
IVG	Infinite verb group	[IVG to_TO be_VB0 IVG]